Small-world networks and the conformation space of a lattice polymer chain

Antonio Scala, Luís A. Nunes Amaral, and Marc Barthélémy
Center for Polymer Studies and Department of Physics, Boston University, MA 02215

We map the conformation space of a simple lattice polymer chain to a network, where (i) the vertices of the network have a one-to-one correspondence to the conformations of the chain, and (ii) a link between two vertices indicates the possibility of switching from one conformation to the other by a single Monte Carlo move of the chain. We find that the geometric properties of this network are similar to those of small-world networks, namely, the diameter of conformation space increases, for large networks, as the logarithm of the number of conformations, while locally the network appears to have low dimensionality.

The physical properties of polymers are the focus of a lot of attention due to (i) their role in many new advanced materials with important technologic applications [1], and (ii) their role in biological processes [2]. An especially important unsolved problem in polymer studies is protein folding. Solving this problem has recently become more urgent as several studies suggest that a number of human diseases, including Parkinson's, Alzheimer's and British dementia, may be due to the aggregation of misfolded proteins [3,4].

The kinetics of protein folding are controlled by the structure of the free-energy landscape [5]. Theoretical calculations predict that in some cases the barriers in the free-energy landscape are quite small [6]. Hence, the diffusion of a protein's conformation on its energy landscape may be determined mainly by the structure and connectivity of conformation space. For this reason, much work has been done on the structure of conformation space and many models have been proposed, including tree structures, random networks, and ultrametric spaces [7]. In spite of all this work, very little is known about the conformation space of polymers in general and proteins in particular.

Here, we present evidence supporting the hypothesis that the conformation space of a lattice homopolymer chain may be a small-world network. These networks —which appear as the result of randomly replacing a fraction p of the links of a d-dimensional lattice with new random links [8]— interpolate between the two limiting cases of a regular lattice ($p = 0$) and a random graph ($p = 1$). The small-world regime is characterized by the properties (i) that a local neighborhood is preserved —as for regular lattices [8]— and (ii) that the average shortest distance between two vertices increases logarithmically with the number of vertices n of the network —as for random graphs [9].

To gain some insight on the conformation space of real polymers, we consider here the conformation space of a 2-dimensional lattice polymer chain. To study the geometrical properties of conformation space, we map it onto a network [10]. We first enumerate *all* allowed conformations of the chain. We then identify (i) each conformation of the chain with a vertex of the network, and (ii) the possibility of changing from one conformation to another, through a *single* Monte Carlo move of the chain, with the existence of a link between the corresponding vertices (Fig. 1).

Our study relies on two important simplifications. The first simplification is to use a lattice model. On this regard, note that due to the limited number of allowed equilibrium angles between two monomers, the modeling of a polymer by a linear chain on a lattice may be experimentally justified [11]. In fact, even for the study of proteins, lattice models led to important insights [10,11]. The second simplification is to neglect interactions. On this regard, note that here we are only interested in determining the *geometric and structural* properties of conformation space, and that interactions affect the *rate of transfer* between the allowed conformations [4,10,12]. Moreover, our results will also apply to any polymer chain in the limit of very high temperature for which the monomer interactions are mostly due to steric effects [13].

We first calculate the geometric properties of conformation space for the simplified lattice polymer model and compare these properties with the predictions of several models for conformation space. This approach is *different* from the ones considered so far in the literature. We do not start by postulating a particular type of structure with some theoretically-desirable properties but, instead, try to generate the full network describing conformation space and compare its properties with different models.

As a first test, we study the dependence of the average shortest distance ℓ between any two vertices in the network on the number n of vertices. Note (i) that the size of the network equals the number of allowed conformations of the chain, and (ii) that the distance between two vertices equals the minimum number of elementary moves of the chain necessary to switch between the two corresponding conformations. For a tree structure or a random network we expect a logarithmic increase of ℓ with n, while for a d-dimensional lattice we expect an algebraic increase: $\ell \sim n^{1/d}$. In contrast, for a small-world network, ℓ follows the scaling law [14]

$$\ell(n,p) \sim (n^*)^{1/d} f(n/n^*)\,, \qquad (1)$$

where the scaling function $f(u)$ has the limits $f(u) \sim$

$u^{1/d}$ for $u \ll 1$ and $f(u) \sim \ln u$ for $u \gg 1$; $n^* \sim p^{-1}$ is a crossover size that separates the large- and small-world regimes, and p is the fraction of "rewired" links [14]. Figure 2 displays our results for the polymer chain, which suggest that the conformation space can be described by a small-world network with $p \approx 10^{-3}$. Figure 2 clearly rules out the possibility that, for $n \gg 1$, the conformation space is a low dimensionality lattice.

As a second test, we study the local structure of conformation space and compare it with that of a random network [8]. To this end, we calculate the clustering coefficient C, which is defined as the average ratio of the number of existing links between neighbors of a vertex and the maximum number of possible links. For a random network [8], we expect $C \simeq z/n$, where z is the average connectivity of the network. In contrast, small-world networks have values of C of the same order of magnitude as those of regular lattices, because only a small percentage of links are different from those in the lattice [8]. In Fig. 3(a), we compare the values of C obtained for the networks with the values of C for random networks with the same size and connectivity. Clearly, the measured clustering coefficients are much larger than the expected values for random networks, ruling out a purely random structure for conformation space.

As a third test, we calculate the number of elementary loops in conformation space —usually referred to as the cyclomatic number [15]— and compare the results with those for a tree structure. For a tree structure the cyclomatic number is identically zero while for all other networks it increases linearly with n. Figure 3(b) shows that the cyclomatic number for the polymer conformation space clearly increases with n, consistent with a small-world network but ruling out a tree structure for conformation space.

As a final test, we calculate the percentage of triplets $\{A, B, C\}$ of vertices in conformation space whose distances obey an ultrametric relation [7]: $d_{AC} \leq \max(d_{AB}, d_{BC})$. Figure 3(c) shows the percentage of ultrametric triplets for conformation space. The percentage of ultrametric triplets is significantly smaller than 100%, ruling out a purely ultrametric structure for conformation space. Moreover, the measured percentage of ultrametric triplets also rules out the random network and the tree structure as descriptions of conformation space.

In summary, the regular lattice is rejected by the first test, the tree structure is rejected by the first, third, and fourth tests, and the random network is rejected by the first, second and fourth tests. Hence, we conclude that the geometrical properties of the conformation space of a lattice polymer chain are consistent with those of small-world networks but not with the geometric properties of the other geometries discussed in the literature.

Next, we address the implications of our finding that the conformation space of a lattice polymer chain may be a small-world network. A central problem in protein folding is the characterization of relaxation processes. The kinetics of a protein's conformations can be mapped to a random walk on conformation space. Naturally, the model studied here is too simplistic to enable us to understand protein folding, however, the study of diffusion on conformation spaces with small-world geometries may still develop our understanding of relaxation in protein folding. Hence, we consider next the problem of diffusion on a small-world network. The usual way to solve a diffusion problem is to obtain the density of states $\rho(\lambda)$ of the transition matrix of the system [16]. For a small-world network, one has [16,17]

$$\rho(\lambda) \sim \lambda^{d/2-1} \ \exp(-p/\sqrt{\lambda}) \,, \tag{2}$$

where the λ's are the eigenvalues of the transition matrix for the network. To quantify the relaxation properties, we calculate the probability of return to the origin, usually considered in the study of random walks [18], which is the Laplace transform of (2),

$$P_0(t) = \int_{1/\ell^2}^{\infty} d\lambda \ \rho(\lambda) \ \exp(-\lambda t) \,, \tag{3}$$

where the lower limit in the integral is a cut-off that takes into account the finite size of the network. The return probability has the scaling behavior

$$P_0(t) - P_0(\infty) \sim \begin{cases} t^{-d/2} & t \ll t_1 \\ \exp(-(p^2 t)^{1/3}) & t_1 \ll t \ll t_2 \\ \exp(-t/\ell^2) & t \gg t_2 \end{cases} , \tag{4}$$

where $P_0(\infty) = 1/n$ and the two crossover times scales can be written, in the small-world regime, as

$$\begin{cases} t_1 \sim 1/p^2 \sim (n^*)^2 \\ t_2 \sim p\ell^3 \sim (n^*)^2 (\ln n)^3 \end{cases} , \tag{5}$$

where we used the result $\ell \sim n^* \ln n$ [14]. The results (2-5) are consistent with our numerical simulations both for small-world networks and for the conformation space of the lattice polymer chain.

The description of the conformation space as a small-world network may be useful for other complex disordered systems such as spin-glasses [19]. Specifically, small-world networks combine the features of apparent *infinite dimensionality* and low connectivity thought to be important for glassy relaxation [19,20]. It has been hypothesized, that in order to reproduce stretched exponential decays [12,21], the space of accessible conformations must have a tree-like structure [16,20]. We have shown that for conformation spaces with a small-world structure stretched exponential relaxation can appear as an intrinsic feature of the space's geometry.

[1] P. Baldus, M. Jansen, and D. Sporn, Science **285**, 699 (1999); A. Hellemans, Science **283**, 771 (1999); R.F. Service, Science **280**, 1691 (1998); A. Garito, R.F. Shi, and M. Wu, Physics Today **47(5)**, 51 (1994).

[2] H. Frauenfelder, P.G. Wolynes, and R.H. Austin, Rev. Mod. Phys. **71**, S419 (1999); R.H. Austin, J.P. Brody, E.C. Cox, T. Duke, and W. Volkmuth, Physics Today **50(2)**, 32 (1997).

[3] P.T. Lansbury, Proc. Nat. Acad. Sci. USA **96**, 3342 (1999).

[4] W.A. Eaton, Proc. Nat. Acad. Sci. USA **96**, 5897 (1999).

[5] J.D. Bryngelson, J.N. Onuchic, N.D. Socci, and P.G. Wolynes, Proteins **21**, 167 (1995).

[6] D. Thirumalai, J. Phys. (France) I **5**, 1457 (1995); J.J. Portman, S. Takada, and P.G. Wolynes, Phys. Rev. Lett. **81**, 5237 (1998).

[7] A.T. Ogielski and D.L. Stein, Phys. Rev. Lett. **55**, 1634 (1985); B. Velikson, T. Garel, H. Orland, and J.C. Smith, J. Comp. Chem. **13**, 1216 (1992); B. Velikson, J. Bascle, T. Garel, and H. Orland, Macromolecules **26**, 4791 (1993).

[8] D.J. Watts and D.H. Strogatz, Nature **393**, 440 (1998); D.J. Watts, *Small Worlds: The Dynamics of Networks Between Order and Randomness (Princeton Reviews in Complexity)* (Princeton University Press, Princeton, New Jersey, 1999).

[9] B. Bollobás, *Random Graphs* (Academic Press, London, 1985).

[10] R. Du, A.Yu. Grosberg, and T. Tanaka, cond-mat/9905106.

[11] H.S. Chan and K.A. Dill, Physics Today **46(2)**, 24 (1993); V. Abkevich, A.M. Gutin, and E.I. Shakhnovich, Biochemistry **33**, 10026 (1994); I. Shrivastava, S Vishveshwara, M. Cieplak, A. Maritan, and J.R. Banavar, Proc. Nat. Acad. Sci. USA **92**, 9206 (1995); V.S. Pande, A.Yu. Grosberg, and T. Tanaka, Rev. Mod. Phys. **72**, 259 (1999); D.K. Klimov and D. Thirumalai, Proc. Nat. Acad. Sci. USA **96**, 6166 (1999).

[12] S.J. Hagen and W.A. Eaton, J. Chem. Phys. **104**, 3395 (1996); J. Sabelko, J. Ervin, and M. Gruebele, Proc. Nat. Acad. Sci. USA **96**, 6031 (1999).

[13] E. Pitard and H. Orland, Europhys. Lett. **41**, 467 (1998).

[14] M. Barthélémy and L.A.N. Amaral, Phys. Rev. Lett. **82**, 3180 (1999); *ibid*, 5180 (1999).

[15] C. Bergé, *Graphs and Hypergraphs* (North-Holland, Amsterdam, 1976) 2nd ed.

[16] A.J. Bray and G.J. Rodgers, Phys. Rev. B **38**, 11461 (1988).

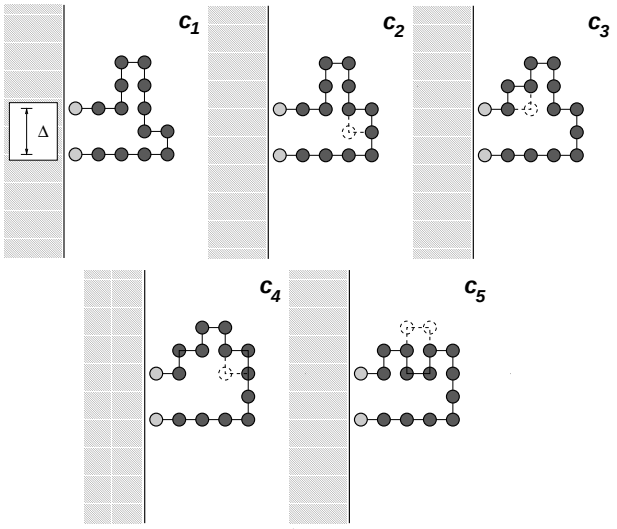
[17] R. Monasson, Europ. Phys. J. B **12**, 555 (1999).

[18] S. Havlin and D. Ben-Avraham, Adv. Phys. **36**, 695 (1987).

[19] D.J. Wales, M.A. Miller, and T.R. Walsh, Nature **394**, 758 (1998); S. Sastry, P.G. Debenedetti, and F.H. Stillinger, Nature **393**, 554 (1998); G. Franzese and A. Coniglio, Phys. Rev. E **59**, 6409 (1999).

[20] I.A. Campbell and L. Bernardi, Phys. Rev. B **50**, 12643 (1994).

[21] M. Lim, T.A. Jackson, and P.A. Anfinrud, Proc. Nat. Acad. Sci. USA **90**, 5801 (1993); J.N. Onuchic, P.G. Wolynes, A. Luthey-Schulten, and N.D. Socci, Proc. Nat. Acad. Sci. USA **92**, 3626 (1995).



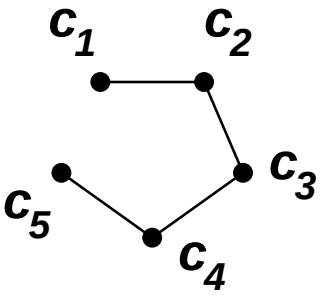


FIG. 1. The conformations of a lattice polymer chain and the mapping of the conformation space to a network. **(a)** The five panels labeled c_1 to c_5 show 5 different conformations of a 15-mer constrained to have the end monomers fixed a distance Δ apart. We use circles with a light shade to represent the fixed monomers, and dark circles to represent the moving monomers. For numerical convenience, we constrain the chain to occupy the half plane bounded by the thick black line. Note that the same results are obtained for other boundary conditions. The conformation of the chain evolves by the usual Monte Carlo elementary moves: the corner flip and the "crankshaft". For example, the chain can switch from conformations c_1 to c_2, c_2 to c_3, and c_3 to c_4 by single corner flips. In panels c_2 to c_5, we use dashed lines and dashed circles to represent the position of the monomer moved from the previous conformation. We use conformations c_4 and c_5 to illustrate the "crankshaft" move, which involves the simultaneous movement of two monomers. **(b)** Mapping of the conformation space of a chain to a network. We first allocate a vertex of the network to each allowed conformation of the chain. We then create a link between two vertices if the two corresponding conformations differ by a single elementary Monte Carlo move. For example, since we can switch between conformations c_1 and c_2 by a single move, we put a link between the vertices c_1 and c_2. On the other hand, because c_5 cannot be reached from c_1 through a single elementary move, we do not place a link between the corresponding vertices.

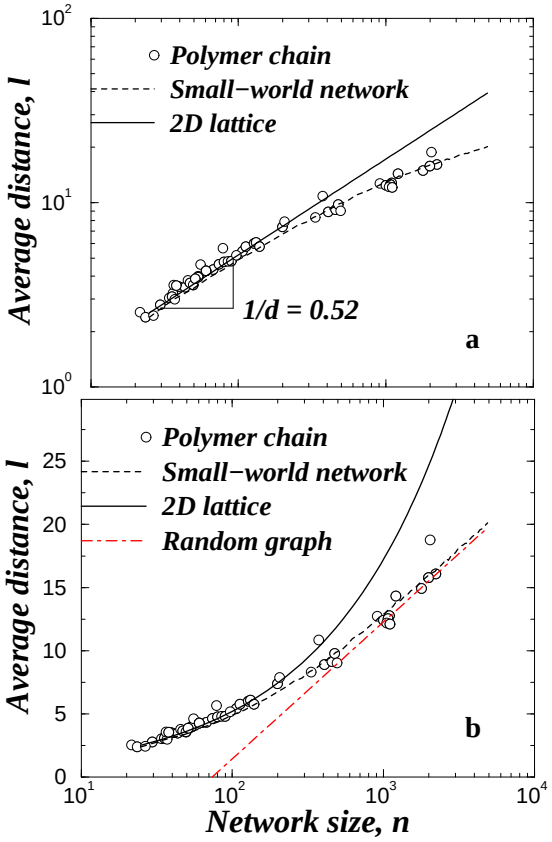
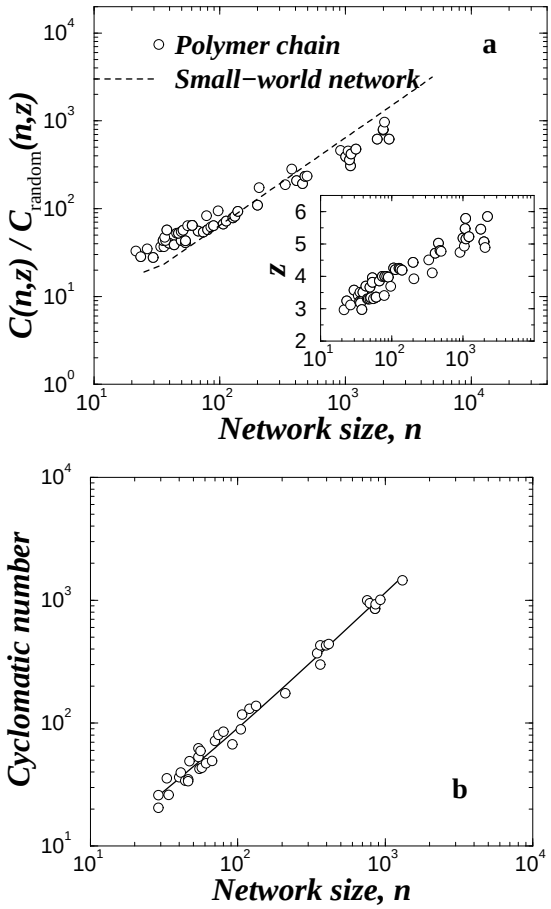


FIG. 2. Average shortest distance between chain conformations. We generate all the possible conformations of the polymer chain defined in Fig. 1 for $\Delta = 1, \ldots, 6$ and for chains with $m = 6, \ldots, 15$ monomers. For each pair of values (Δ, m), we identify in the conformation space all the different *connected networks* and calculate their size n. Here, we calculate the average shortest distance between every pair of vertices using to the breadth-first search algorithm. **(a)** Loglog plot of the average distance, rescaled according to the value of average connectivity [8]. For $n < 100$, the average distance ℓ increases as $n^{1/2}$, as it would for a 2-dimensional lattice. **(b)** Same data but in log-linear plot. For $n > 1000$, we observe $\ell \sim \ln n$, as for a random network, and as predicted by Eq. (1). Our results are consistent with the case of a small-world network with $p \approx 10^{-3}$, clearly ruling out a low dimensional lattice as a model of conformation space. Note that for this and all other figures, the results for the small-world network have been average over 100 realizations of the network while each symbol "○" corresponds to a single conformation network. Hence, there is far less noise for the small-world networks data.



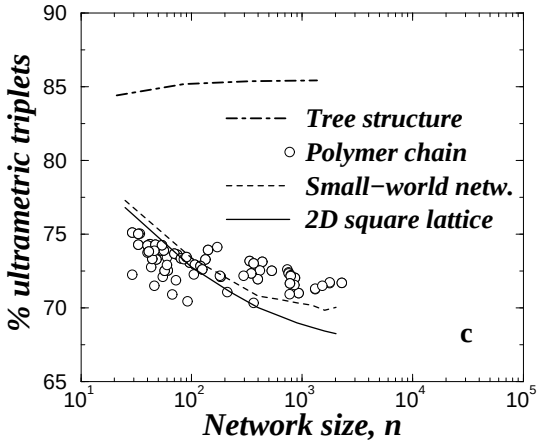


FIG. 3. Structure of conformation space. **(a)** Normalized clustering coefficients for the polymer chain, $C(n,z)/C_{\rm random}(n,z) \simeq nC(n,z)/z$, where z is the average connectivity. For a small-world network, the clustering coefficient will be approximately a constant (for fixed p), and the normalized clustering coefficient increases linearly with n. Our results clearly rules out a purely random structure for conformation space since the normalized clustering coefficient for the polymer chain conformation space is orders of magnitude larger than the value for a random network. Note that the apparently logarithmic discrepancy between the two curves in the figure may be due to the logarithmic increase of z with n, which is shown in the inset. **(b)** Cyclomatic number. The cyclomatic number for a tree structure is identically zero because there are no loops. For a generic network [15], the cyclomatic number increases as nz. Hence, we plot the cyclomatic number for the polymer conformation space normalized by the connectivity z. As expected, we observe a linear increase with size n. **(c)** Percentage of ultrametric triplets. We calculate the number of triplets obeying the relation $d_{AC} \leq \max(d_{AB}, d_{BC})$. For a purely ultrametric space, all triplets have distances obeying this relation. For 2-dimensional square lattices and for small-world networks only slightly more than 2/3 of the triplets —which is the lower bound— obey the ultrametric relation. In contrast, for random networks and tree structures, we find a higher percentage of triplets to be ultrametric. It is apparent that values for the random network and tree structure are not consistent with the measured values for the polymer conformation space. Note that each data point for the polymer chain corresponds to a single conformation network while the results for the small-world network represent an average over 100 networks.